THAT WHICH IS CLAIMED IS:

1. A method of distributing workload between data processing systems executing an application which communicates over a network, the method comprising:

receiving a request for a connection to the application over the network;

obtaining workload information for the data processing systems;

obtaining network quality of service information associated with communications over the network for respective ones of the data processing systems;

generating workload metrics associated with respective ones of the data processing systems utilizing the workload information and the corresponding network quality of service information for the data processing systems; and

distributing the requested connection to instances of the application executing on the data processing systems based on the generated workload metrics.

2. A method according to Claim 1, wherein the network quality of service information comprises at least one of network packet loss information, network timeout information, and number of connections information.

3. A method according to Claim 1, wherein the workload information comprises a weight value (W) corresponding to a data processing systems processing capacity.

4.   A method according to Claim 3, wherein the network quality of service information comprises a loss weight (F-loss) based on packet loss ratio, a network timeout weight (F-timeout) based on a network timeout ratio and a connection weight (F-con) based on a number of active connections.

5.   A method according to Claim 4, wherein F-loss is a ratio of retransmitted packets to total transmitted packets, wherein F-timeout is a ratio of number of timeouts to number of transmitted segments, and wherein F-con is a ratio of current connections to total allowed connections.

6.   A method according to Claim 4, further comprising the steps of:
    determining a value for F-loss by mapping a TCP loss ratio to the value for F-loss;
    determining a value for F-timeout by mapping a timeout loss ratio to the value for F-timeout; and
    determining a value for F-conn by determining if a number of current connections exceed a predefined percentage of a total number of allowed connections and setting F-con to 1 if the total number of connections exceeds the predefined percentage of the total number of allowed connections and to 0 if the total number of connections does not exceed the predefined percentage of the total number of allowed connections.

7.   A method according to Claim 4, wherein the step of generating workload metrics comprises the step of evaluating the equation:

$$Aw = (1 - MIN(1, F\text{-}loss + F\text{-}timeout + F\text{-}con)) * W$$

where Aw is the workload metric.

8.   A method according to Claim 1, wherein the quality of service information comprises quality of service information for an instance of the application executing on one of the data processing systems.

9.   A method according to Claim 1, wherein the quality of service information is divided into classes of quality of service information associated with the application based on a common constraint, the method further comprising the step of:

determining a class of quality of service information associated with the connection request; and

wherein the step of obtaining network quality of service information associated with communications over the network for the data processing systems comprises obtaining network quality of service information for the class of quality of service information associated with the connection request.

10.   A method according to Claim 9, wherein the step of generating workload metrics comprises the step of combining the workload information and the corresponding network quality of service information for the data processing systems based on the class of quality of service information associated with the request so as to

provide workload metrics based on the workload information, the quality of service information and the class of quality of service information associated with the request.

11. A method according to Claim 1, wherein the data processing systems comprise data processing systems in a Sysplex, wherein the steps of receiving a request for a connection to the application over the network, obtaining workload information for the data processing systems, obtaining network quality of service information associated with communications over the network for the data processing systems, generating workload metrics and distributing the requested connection to application instances on the data processing systems based on the workload metric are carried out by a routing communication protocol stack in the Sysplex.

12. A method according to Claim 11 wherein the routing communication protocol stack obtains the quality of service information from policy agents executing on the data processing systems in the Sysplex.

13. A system for workload distribution, comprising:

a workload distributer which selects data processing systems in a cluster of data processing systems for distribution of connections based on quality of service information associated with the data processing systems and workload information associated with the data processing systems; and

a router operably associated with the workload distributer which receives requests for connection to an application executing on ones of the data processing systems and distributes the connections to data

5   processing systems in the cluster of data processing systems selected by the workload distributer.

14. A system according to Claim 13, wherein the router comprises a routing communication protocol stack.

10

15. A system according to Claim 14, wherein the requests for connections to the application comprise requests for connections to a dynamically routable virtual Internet Protocol address.

15

16. A system according to Claim 13, further comprising policy agents associated with the data processing systems which provide quality of service information to the workload distributer.

20

17. A system according to Claim 13, wherein the network quality of service information comprises a loss weight (F-loss) based on packet loss ratio, a network timeout weight (F-timeout) based on a network timeout

25   ratio and a connection weight (F-con) based on a number of active connections.

18. A system according to Claim 17, wherein the workload distributer is configured to evaluate the

30   equation:

Aw = (1 - MIN(1, F-loss + F-timeout + F-con)) * W

where Aw is a workload metric and W is the workload information and wherein the workload distributer selects a data processing system with a best Aw value.

5   19.   A system according to Claim 17, wherein F-loss is determined by mapping a TCP loss ratio to the value for F-loss, F-timeout is determined by mapping a timeout loss ratio to the value for F-timeout and F-conn has a value of 1 if a total number of connections to the

10   application associated with the connection request exceeds a predefined percentage of the total number of allowed connections and to 0 if the total number of connections does not exceed the predefined percentage of the total number of allowed connections.

15
     20.   A system according to Claim 18, wherein the workload distributer is configured to evaluate the equation:

$$Aw = (1 - MIN(1, F\text{-}loss + F\text{-}timeout + F\text{-}con)) * W$$

20   where Aw is a workload metric and W is the workload information and wherein the workload distributer selects a data processing system with a best Aw value.

     21.   A system for distributing workload between data

25   processing systems executing an application which communicates over a network, comprising:
     means for receiving a request for a connection to the application over the network;
     means for obtaining workload information for the

30   data processing systems;

means for obtaining network quality of service information associated with communications over the network for respective ones of the data processing systems;

5   means for generating workload metrics associated with respective ones of the data processing systems utilizing the workload information and the corresponding network quality of service information for the data processing systems; and

10   means for distributing the requested connection to instances of the application executing on the data processing systems based on the generated workload metrics.

15   22.   A system according to Claim 21, wherein the network quality of service information comprises at least one of network packet loss information, network timeout information, and number of connections information.

20   23.   A system according to Claim 21, wherein the workload information comprises a weight value (W) corresponding to a data processing systems processing capacity.

25   24.   A system according to Claim 23, wherein the network quality of service information comprises a loss weight (F-loss) based on packet loss ratio, a network timeout weight (F-timeout) based on a network timeout ratio and a connection weight (F-con) based on a number

30   of active connections.

25.   A system according to Claim 24, wherein F-loss
is a ratio of retransmitted packets to total transmitted
packets, wherein F-timeout is a ratio of number of
timeouts to number of transmitted segments, and wherein
5   F-con is a ratio of current connections to total allowed
connections.

26.   A system according to Claim 24, further
comprising:
10   means for determining a value for F-loss by mapping
a TCP loss ratio to the value for F-loss;
        means for determining a value for F-timeout by
mapping a timeout loss ratio to the value for F-timeout;
and
15   means for determining a value for F-conn by
determining if a number of current connections exceed a
predefined percentage of a total number of allowed
connections and setting F-con to 1 if the total number of
connections exceeds the predefined percentage of the
20   total number of allowed connections and to 0 if the total
number of connections does not exceed the predefined
percentage of the total number of allowed connections.

27.   A system according to Claim 24, wherein the
25   means for generating workload metrics comprises means for
evaluating the equation:
        Aw = (1 - MIN(1, F-loss + F-timeout + F-con)) * W
where Aw is the workload metric.

30   28.   A system according to Claim 21, wherein the
quality of service information comprises quality of

service information for an instance of the application
executing on one of the data processing systems.

29. A system according to Claim 21, wherein the
5   quality of service information is divided into classes of
quality of service information associated with the
application based on a common constraint, further
comprising:
    means for determining a class of quality of service
10   information associated with the connection request; and
    wherein the means for obtaining network quality of
service information associated with communications over
the network for the data processing systems comprises
means for obtaining network quality of service
15   information for the class of quality of service
information associated with the connection request.

30. A system according to Claim 29, wherein the
means for generating workload metrics comprises means for
20   combining the workload information and the corresponding
network quality of service information for the data
processing systems based on the class of quality of
service information associated with the request so as to
provide workload metrics based on the workload
25   information, the quality of service information and the
class of quality of service information associated with
the request.

31. A computer program product for distributing
30   workload between data processing systems executing an

application which communicates over a network, comprising:

a computer readable storage medium having computer readable program code embodied therein, the computer readable program code comprising:

computer readable program code which receives a request for a connection to the application over the network;

computer readable program code which obtains workload information for the data processing systems;

computer readable program code which obtains network quality of service information associated with communications over the network for respective ones of the data processing systems;

computer readable program code which generates workload metrics associated with respective ones of the data processing systems utilizing the workload information and the corresponding network quality of service information for the data processing systems; and

computer readable program code which distributes the requested connection to instances of the application executing on the data processing systems based on the generated workload metrics.

32. A computer program product according to Claim 31, wherein the network quality of service information comprises at least one of network packet loss information, network timeout information, and number of connections information.

33. A computer program product according to Claim 31, wherein the workload information comprises a weight value (W) corresponding to a data processing systems processing capacity.

34. A computer program product according to Claim 33, wherein the network quality of service information comprises a loss weight (F-loss) based on packet loss ratio, a network timeout weight (F-timeout) based on a network timeout ratio and a connection weight (F-con) based on a number of active connections.

35. A computer program product according to Claim 34, wherein F-loss is a ratio of retransmitted packets to total transmitted packets, wherein F-timeout is a ratio of number of timeouts to number of transmitted segments, and wherein F-con is a ratio of current connections to total allowed connections.

36. A computer program product according to Claim 34, further comprising:

computer readable program code which determines a value for F-loss by mapping a TCP loss ratio to the value for F-loss;

computer readable program code which determines a value for F-timeout by mapping a timeout loss ratio to the value for F-timeout; and

computer readable program code which determines a value for F-conn by determining if a number of current connections exceed a predefined percentage of a total number of allowed connections and setting F-con to 1 if

the total number of connections exceeds the predefined percentage of the total number of allowed connections and to 0 if the total number of connections does not exceed the predefined percentage of the total number of allowed connections.

37. A computer program product according to Claim 34, wherein the computer readable program code which generates workload metrics comprises computer readable program code which evaluates the equation:

$$Aw = (1 - MIN(1, F\text{-loss} + F\text{-timeout} + F\text{-con})) * W$$

where Aw is the workload metric.

38. A computer program product according to Claim 31, wherein the quality of service information comprises quality of service information for an instance of the application executing on one of the data processing systems.

39. A computer program product according to Claim 31, wherein the quality of service information is divided into classes of quality of service information associated with the application based on a common constraint, further comprising:

computer readable program code which determines a class of quality of service information associated with the connection request; and

wherein the computer readable program code which obtains network quality of service information associated with communications over the network for the data processing systems comprises computer readable program

code which obtains network quality of service information for the class of quality of service information associated with the connection request.

5   40. A computer program product according to Claim 39, wherein the computer readable program code which generates workload metrics comprises computer readable program code which combines the workload information and the corresponding network quality of service information
10   for the data processing systems based on the class of quality of service information associated with the request so as to provide workload metrics based on the workload information, the quality of service information and the class of quality of service information
15   associated with the request.